ORIENTATION DETERMINATION FROM CRYO-EM IMAGES USING LEAST UNSQUARED DEVIATION

LANHUI WANG*, AMIT SINGER†, AND ZAIWEN WEN‡

Abstract. A major challenge in single particle reconstruction from cryo electron-microscopy is to establish a reliable ab-initio 3D model using 2D images with unknown orientations. Common-lines based methods can determine the orientations of images without additional geometric information. However, such methods fail when the detection rate of common-lines is too low due to the high level of noise in the images. An approximation to the least squares global self consistency error was obtained in [29] using convex relaxation by semidefinite programming. The purpose of this paper is three-fold. First, we introduce another, yet more robust, global self consistency error in an optimization problem that can be solved via semidefinite relaxation. Second, we introduce a spectral norm constraint to robustify the relaxed problem. Third, we use alternating direction method or the iteratively reweighted least squares (IRLS) procedure to solve the problem. Numerical experiments demonstrate that the proposed method significantly decreases the orientation estimation error when the detection rate of common-lines is low.

1. Introduction. In single particle analysis, cryo-electron microscopy (cryo-EM) is used to attain a resolution sufficient to interpret fine details in three-dimensional (3D) macromolecular structures [9, 37, 10, 38]. Cryo-EM is used to acquire 2D projection images of thousands of individual, identical frozen-hydrated macromolecules at random unknown orientations and positions. The collected images are extremely noisy due to the limited electron dose used for imaging to avoid excessive beam damage. In addition, the unknown orientational information of the imaged particles need to be estimated for 3D reconstruction. An ab-initio estimation of the orientations of images using the random-conical tilt technique [24] or common-lines based approaches [35, 28, 29] are often applied after multivariate statistical analysis [13, 36] and classification techniques [34, 21, 30] that are used to sort and partition the large set of images by their viewing directions, producing "class averages" of enhanced signal-to-noise ratio (SNR). Using the ab-initio estimation of the orientations, a preliminary 3D map is reconstructed from the images by a 3D reconstruction algorithm. The initial model is then iteratively refined [20] in order to obtain a higher-resolution 3D reconstruction.

The Fourier projection-slice theorem (page 11 in [18]) plays a fundamental role in the common-lines based reconstruction methods. The theorem states that a slice extracted from the frequency domain representation of a 3D volume yields the Fourier transform of a 2D projection of the volume in a direction perpendicular to the slice (Figure 1.1). Thus, any two projections imaged from different viewing angles will intersect at a line in Fourier space, which is called the common-line between the two images. The common-lines between any three images with different projection directions determine their relative orientation up to handedness. This is the basis of the "angular reconstitution" technique of van Heel [35], which was also developed independently by Vainshtein and Goncharov [33]. This technique uses a brute force

*The Program in Applied and Computational Mathematics (PACM), Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA, lanhuiw@math.princeton.edu, Corresponding author. Tel.: +1 609 258 5785; fax: +1 609 258 1735.

†Department of Mathematics and PACM, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, USA, amits@math.princeton.edu

‡Department of Mathematics and Institute of Natural Sciences, Shanghai Jiao Tong University, 512, Pao Yue-Kong Library, 800 Dongchuan Rd, Shanghai Jiao Tong University, China, zw2109@sjtu.edu.cn

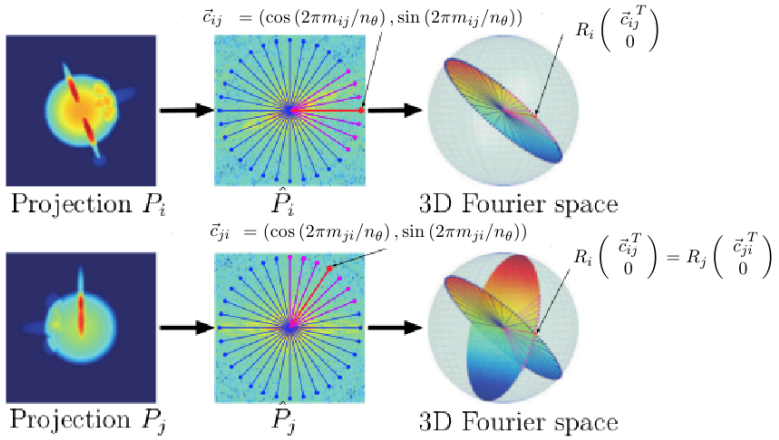


Fig. 1.1: Fourier projection-slice theorem. In the middle, $\hat{P}_i$ is a polar Fourier transform of projection P_i on the left. The red line $\vec{c}_{ij}$ represents the direction of a common-line between $\hat{P}_i$ and $\hat{P}_j$ on $\hat{P}_i$. On the right, the two transformed images $\hat{P}_i$ and $\hat{P}_j$ intersect with each other at the common-line after rotations R_i and R_j, yielding the equation (3.3).

approach to include additional projections. Farrow and Ottensmeyer [7] used quaternions to obtain the relative orientation of a new projection in a least square sense. The main problem with such techniques is that they are sensitive to false detection of common lines that leads to the accumulation of errors. Penczek et.al. [23] tried to obtain the rotations corresponding to all projections simultaneously by minimizing a global energy functional, which requires a brute force search in an exponentially large parametric space of all possible orientations for all projections. Mallick et. al. [16] and Singer et al. [28] applied Bayesian approaches to use common-lines information from different groups of projections. Recently, Singer et. al. [29] developed two algorithms based on eigenvectors and semidefinite programming for estimating the orientations of all images. These correspond to convex relaxations of the global self-consistency error minimization. The two algorithms accurately estimate all orientations at relatively low common-line detection rates.

When the signal-to-noise ratio (SNR) of the image is significantly low, the detected common-lines consist of a modest number of noisy inliers, which are explained well by the image orientations, along with a large number of outliers, that have no structure. The standard common-lines based methods, including those using the least squares (LS) [7, 29], are sensitive to these outliers. Here we propose to estimate the orientations by minimizing a different, more robust self consistency error, which is the sum of unsquared residuals [19, 32], rather than the sum of squared residuals in the LS formulation. The minimization problem is solved via semidefinite relaxation. When the detection rate of common-lines is extremely low, a spectral norm constraint is added to robustify the result. The minimization problem can be either solved by alternating direction method (ADM) or the iteratively reweighted LS (IRLS) procedure.

2. Detection of common-lines between images. Typically, the first step for detecting common lines is to compute the 2D Fourier transform of each image on a polar grid using, e.g., the non-uniform fast Fourier transform (NUFFT) [6, 8, 12]. The transformed images have resolution n_r in the radial direction and resolution n_θ in the angular direction, that is, the radial resolution n_r is the number of equi-spaced samples along each ray in the radial direction, and the angular resolution n_θ is the number of angularly equis-spaced Fourier rays computed for each image (Figure 1.1). For simplicity, we let n_θ be an even number. The transformed images are denoted as $\left(\vec{l}_0^k, \vec{l}_1^k, \ldots, \vec{l}_{n_\theta-1}^k\right)$, where $\vec{l}_m^k = \left(l_{m,1}^k, l_{m,2}^k, \ldots, l_{m,n_r}^k\right)$ is an n_r dimensional vector, $m \in \{0, 1, \ldots, n_\theta - 1\}$ is the index of a ray, $k \in \{1, 2, \ldots, K\}$ is the index of an image and K is the number of images. The DC term is shared by all lines independently of the image, and is therefore excluded for comparison. To determine the common line between two images P_i and P_j, the similarity between all n_θ radial lines $\vec{l}_0^i, \vec{l}_1^i, \ldots, \vec{l}_{n_\theta-1}^i$ from the first image with all n_θ radial lines $\vec{l}_0^j, \vec{l}_1^j, \ldots, \vec{l}_{n_\theta-1}^j$ from the second image are measured (overall n_θ^2 comparisons), and the pair of radial lines $\vec{l}_{m_{i,j}}^i$ and $\vec{l}_{m_{j,i}}^j$ with the highest similarity are identified as the common-line pair between the two images. However, as a radial line is the complex conjugate of its antipodal line, the similarity measure between $\vec{l}_{m_1}^i$ and $\vec{l}_{m_2}^j$ has the same value as that between their antipodal lines $\vec{l}_{m_1+n_\theta/2}^i$ and $\vec{l}_{m_2+n_\theta/2}^j$ (where addition of indices is taken modulo n_θ). Thus the number of distinct similarity measures that need to be computed is $n_\theta^2/2$ obtained by restricting the index m_1 to take values between 0 and $n_\theta/2$ and letting m_2 take any of the n_θ possibilities (see also [35] and [22], p. 255). Equivalently, it is possible to compare real valued 1D line projections of the 2D projection images, instead of comparing radial Fourier lines that are complex valued. According to the Fourier projection-slice theorem, each 1D projection is obtained by the inverse Fourier transform of the corresponding Fourier radial line $\vec{l}_m^k$ and its antipodal line $\vec{l}_{m+n_\theta/2}^k$, and is denoted as $\vec{s}_m^k$. The 1D projection lines of a cryo-EM image can be displayed as a 2D image known as a "sinogram" (see [35, 26]).

Traditionally, the pair of radial lines (or sinogram lines) that has the maximum normalized cross correlation is declared as the common line, that is,

$$(m_{i,j}, m_{j,i}) = \underset{0 \le m_1 \imath n_\theta/2, 0 \le m_2 \imath n_\theta}{\arg\max} \frac{\left\langle \vec{l}_{m_1}^i, \vec{l}_{m_2}^j \right\rangle}{\left\| \vec{l}_{m_1}^i \right\| \left\| \vec{l}_{m_2}^j \right\|}, \text{ for all } i \neq j, \tag{2.1}$$

where $m_{i,j}$ is a discrete estimate for where the j'th image intersects with the i'th image. In practice, a weighted correlation, which is equivalent to applying a combination of high-pass and low-pass filters is used to determine proximity. As noted in [35], the normalization is performed so that the correlation coefficient becomes a more reliable measure of similarity between radial lines. Note that even with clean images, this estimate will have a small deviation from its ground truth (unknown) value due to discretization errors. With noisy images, large deviations of the estimates from their true values (say, errors of more than 10°) are frequent, and their frequency increases with the level of noise. We refer to common lines whose $m_{i,j}$ and $m_{j,i}$ values were estimated accurately (up to a given discretization error tolerance) as "correctly detected" common lines, or "inliers" and to the remaining common lines as "falsely detected", or "outliers".

3. The weighted LS and the least unsquared deviation (LUD). We define the directions of detected common-lines between the transformed image i and

transformed image j as unit vectors (Figure 1.1)

$$\vec{c}_{ij} = \left(c_{ij}^1, c_{ij}^2\right) = \left(\cos\left(2\pi m_{ij}/n_\theta\right), \sin\left(2\pi m_{ij}/n_\theta\right)\right), \tag{3.1}$$

$$\vec{c}_{ji} = \left(c_{ji}^1, c_{ji}^2\right) = \left(\cos\left(2\pi m_{ji}/n_\theta\right), \sin\left(2\pi m_{ji}/n_\theta\right)\right), \tag{3.2}$$

where $\vec{c}_{ij}$ and $\vec{c}_{ji}$ are on the transformed images i and j respectively, and m_{ij} and m_{ji} are discrete estimate for the common lines' positions using (2.1). Let the rotation matrices $R_i \in \mathbf{SO}(3)$, $i = 1, \cdots, K$ represent the orientations of the K images. According to the Fourier projection-slice theorem, the common lines on every two images should be the same after the 2D transformed images are inserted in the 3D Fourier space using the corresponding rotation matrices, that is,

$$R_i \begin{pmatrix} \vec{c}_{ij}^T \\ 0 \end{pmatrix} = R_j \begin{pmatrix} \vec{c}_{ji}^T \\ 0 \end{pmatrix} \text{ for } 1 \le i < j \le K. \tag{3.3}$$

These can be viewed as $\begin{pmatrix} K \\ 2 \end{pmatrix}$ linear equations for the $6K$ variables corresponding to the first two columns of the rotation matrices (the third column of each rotation matrix does not contribute in (3.3) due to the zero third entries in the common-line vectors in $\mathbb{R}^3$). The weighted LS approach for solving this system can be formulated as the minimization problem

$$\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} w_{ij} \left\| R_i \left(\vec{c}_{ij}, 0\right)^T - R_j \left(\vec{c}_{ji}, 0\right)^T \right\|^2, \tag{3.4}$$

where the weights w_{ij} indicates the confidence in the detections of common-lines between pairs of images. Since $\left(\vec{c}_{ij}, 0\right)^T$ and $\left(\vec{c}_{ji}, 0\right)^T$ are 3D unit vectors, their rotations are also unit vectors; that is, $\left\| R_i \left(\vec{c}_{ij}, 0\right)^T \right\| = \left\| R_j \left(\vec{c}_{ji}, 0\right)^T \right\| = 1$. It follows that the minimization problem (3.4) is equivalent to the maximization problem of the sum of dot products

$$\max_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} w_{ij} \langle R_i \left(\vec{c}_{ij}, 0\right)^T, R_j \left(\vec{c}_{ji}, 0\right)^T \rangle \tag{3.5}$$

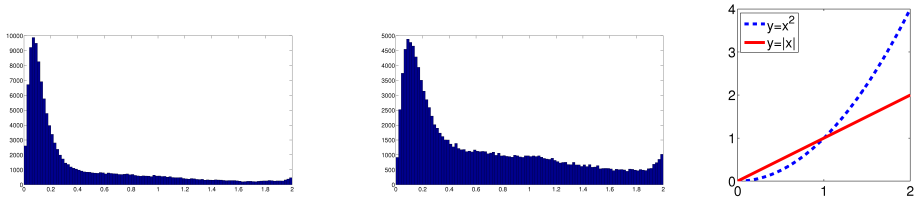
subject to the constraints (4.1). When the weight $w_{ij} = 1$ for each pair $i \neq j$, (3.5) is equivalent to the LS problem. The LS approach was considered in [23], and more recently in [29] using convex relaxation of the non-convex constraint set. The least squares approach may not be optimal however in our case due to the typically large proportion of outliers (Figure 3.1).

To guard the orientation estimations against outliers, we replace the sum of weighted squared residuals in (3.4) with the more robust sum of unsquared residuals and obtain

$$\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} \left\| R_i \left(\vec{c}_{ij}, 0\right)^T - R_j \left(\vec{c}_{ji}, 0\right)^T \right\|, \tag{3.6}$$

or equivalently,

$$\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} \left\| \left(\vec{c}_{ij}, 0\right)^T - R_i^T R_j \left(\vec{c}_{ji}, 0\right)^T \right\|, \tag{3.7}$$



(a) SNR=1/32 (b) SNR=1/64 (c) x^2 vs $|x|$

Fig. 3.1: Left and Middle: The histogram plots of errors in the detected common-lines $\vec{c}_{ij}$ for all i and j, i.e., $\left\| R_i \left(\vec{c}_{ij}, 0\right)^T - R_j \left(\vec{c}_{ji}, 0\right)^T \right\|$ where R_i is a true rotation matrix for all i. Right: elucidating the different between the squared distance and the absolute deviation.

and we refer to this problem as least unsquared deviation (LUD) problem. The self consistency error given in (3.6) reduces the contribution from large residuals that may result from outliers (Figure 3.1c) [1].

4. Semidefinite Programming Relaxation (SDR) and the Rounding procedure. However, both the weighted LS problem (3.4) and the LUD problem (3.6) are non-convex and therefore extremely difficult to solve if one requires the matrices R_i to be rotations, that is, when adding the constraints

$$R_i R_i^T = I_3, \ \det\left(R_i\right) = 1 \text{ for } i = 1, \ldots, K, \tag{4.1}$$

where I_3 is the 3×3 identity matrix. A relaxation method that neglects the constraints (4.1) will simply collapse to the trivial solution $R_1 = \ldots = R_K = 0$ which obviously does not satisfy the constraint (4.1).

The relaxation in [29] that uses semidefinite programming (SDP) can be modified in a straightforward manner in order to deal with non-unity weights w_{ij} in (3.5). We present this modification here for two reasons. The first reason is making the exposition as self contained as possible. The second reason is that the rounding procedure after SDP is slightly different than the one presented in [29] and is closer in spirit to the rounding procedure of Goemans and Williamson for the MAX-CUT problem [11].

4.1. Constructing the Gram matrix G from the rotations R_i. We denote the columns of the rotation matrix R_i by R_i^1 , R_i^2 , and R_i^3 , and write the rotation matrices as

$$R_i = \begin{pmatrix} | & | & | \\ R_i^1 & R_i^2 & R_i^3 \\ | & | & | \end{pmatrix}, \ i = 1, \ldots, K.$$

We define a $3 \times 2K$ matrix R given by

$$R = \begin{pmatrix} | & | & & | & | & & | & | \\ R_1^1 & R_1^2 & \cdots & R_k^1 & R_k^2 & \cdots & R_K^1 & R_K^2 \\ | & | & & | & | & & | & | \end{pmatrix}. \tag{4.2}$$

[1] The weighted version of (3.6) is $\min_{R_1,\ldots,R_K \in \mathbf{SO}(3)} \sum_{i \neq j} w_{ij} \left\| R_i \left(\vec{c}_{ij}, 0\right)^T - R_j \left(\vec{c}_{ji}, 0\right)^T \right\|$. We focus on analyzing (3.6) since the analysis for its weighted version is similar

The Gram matrix G for the matrix R is a $2K \times 2K$ matrix of inner products between the 3D column vectors of R, that is,

$$G = R^T R. \tag{4.3}$$

Clearly, G is a rank-3 semidefinite positive matrix ($G \succcurlyeq 0$), which can be conveniently written as a block matrix

$$G = (G_{ij})_{i,j=1,\cdots,K},$$

where G_{ij} is the 2×2 upper left block of the rotation matrix $R_i^T R_j$, that is,

$$G_{ij} = \begin{pmatrix} (R_i^1)^T \\ (R_i^2)^T \end{pmatrix} \left(R_i^1 R_i^2\right).$$

The orthogonality of the rotation matrices ($R_i^T R_i = I$) implies that

$$G_{ii} = I_2, \ i = 1, 2, \cdots, K, \tag{4.4}$$

where I_2 is a 2×2 identity matrix.

4.2. SDR for weighted LS. We first define two $2K \times 2K$ matrices $S = (S_{ij})_{i,j=1,\cdots,K}$ and $W = (W_{ij})_{i,j=1,\cdots,K}$, where the 2×2 sub-blocks S_{ij} and W_{ij} are given by

$$S_{ij} = \vec{c}_{ji}^T \vec{c}_{ij} \text{ and } W_{ij} = w_{ij} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

Both the matrices S and W are symmetric and they stores all available common-line information and weight information respectively. It follows that the objective function (3.5) is the trace of the matrix $(\beta \circ S)\, G$:

$$\sum_{i \neq j} w_{ij} \langle R_i \left(\vec{c}_{ij}, 0\right)^T, R_j \left(\vec{c}_{ji}, 0\right)^T \rangle = \operatorname{trace}\left((W \circ S)\, G\right), \tag{4.5}$$

where the symbol $\circ$ denotes the Hadamard product between two matrices. A natural relaxation of the optimization problem (3.5) is thus given by the SDP

$$\max_{G \in \mathbb{R}^{2K \times 2K}} \operatorname{trace}\left((W \circ S)\, G\right) \tag{4.6}$$

$$\text{s.t.} \quad G_{ii} = I_2, \ i = 1, 2, \cdots, K. \tag{4.7}$$

$$G \succcurlyeq 0 \tag{4.8}$$

The only constraints missing in this SDP formulation is the non-convex rank-3 constraint on the Gram matrix G, which is referred as a semidefinite relaxation (SDR) [15]. The problem (4.6)-(4.8) is a standard SDP problem, and it can be well solved by the solver SDPLR [3] which takes advantage of the low-rank property of G. SDPLR is a first-order algorithm via low-rank factorization and hence can provide approximate solutions for large scale problems. Moreover, the iterations of SDPLR are extremely fast.

4.3. SDR for LUD. Similar to defining the Gram matrix G in (4.3), we define a $3K \times 3K$ matrix $\tilde{G}$ as $\tilde{G} = (\tilde{G}_{ij})_{i,j=1,\cdots,K}$, where the 3×3 blocks $\tilde{G}_{ij} = R_i^T R_j$. Then a natural SDR for (3.7) is

$$\min_{\tilde{G} \succcurlyeq 0} \sum_{i \neq j} \left\| (\vec{c}_{ij}, 0)^T - \tilde{G}_{ij} \, (\vec{c}_{ji}, 0)^T \right\|, \text{ s.t. } \tilde{G}_{ii} = I_3. \tag{4.9}$$

The constraints missing in this SDP formulation are the non-convex rank-3 constraint and the determinant constraints $\det(\tilde{G}_{ij}) = 1$ on the Gram matrix $\tilde{G}$. However, the solution $\tilde{G}$ to (4.9) is not unique. Note that if a set of rotation matrices $\{R_i\}$ is the solution to (3.7), then the set of reflected rotation matrices $\{JR_iJ\}$ is also the solution to (3.7), where the matrix J is defined as

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Thus another solution to (4.9) is the Gram matrix $\tilde{G}^J = (\tilde{G}^J_{ij})_{i,j=1,\cdots,K}$ with the 2×2 sub-blocks given by $\tilde{G}^J_{ij} = JR_i^T JJR_jJ = JR_i^T R_j J$. It can be verified that $\frac{1}{2}(\tilde{G} + \tilde{G}^J)$ is also the solution to (4.9). Using the fact that

$$\frac{1}{2}(\tilde{G}_{ij} + \tilde{G}^J_{ij}) = \begin{pmatrix} & & 0 \\ & G_{ij} & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

the problem (4.9) is reduced to

$$\min_{G \succcurlyeq 0} \sum_{i \neq j} \left\| \vec{c}_{ij}^T - G_{ij} \vec{c}_{ji}^T \right\|, \text{ s.t. } G_{ii} = I_2. \tag{4.10}$$

This is a SDR for the LUD problem (3.6). The problem (4.10) can be solved using alternating direction method (see details in section 6.2).

4.4. The Randomized Rounding Procedure. The matrix R is recovered from a random projection of the solution G of the SDP (4.6). We randomly draw a $2K \times 3$ matrix P from the Stiefel manifold $V_3(\mathbb{R}^{2K})$. The random matrix P is computed using the orthogonal matrix Q and the upper triangular matrix R from QR factorization of a random matrix with standard normal entries, that is, $P = Q \, \text{sign}\,(\text{diag}\,(R))$, where sign is the sign function and $\text{diag}(R)$ is an diagonal matrix whose diagonal entries are the same as those in the matrix R. The matrix P is shown to be drawn uniformly from the Stiefel manifold in [17]. We project the solution G onto the subspace spanned by the three columns of the matrix GP [2]. We therefore expect to be able to recover the first two columns of the rotation matrices $R_1, \ldots, R_K$ from the three column vectors $\vec{u}^1, \vec{u}^2$ and $\vec{u}^3$ of the the first 3 columns of an unitary matrix U, where U is from singular value decomposition (SVD) of the $2K \times 3$ matrix GP. Note that it is possible to recover the molecule only up to a global orthogonal transformation, that is, up to rotation and possibly reflection, since the unitary matrix

[2] The 3 dimensional subspace can also be spanned by the eigenvectors associated to the top three eigenvalues of G. In addition, the fourth largest eigenvalue is expected to be nearly zero. See more details in [29].

U is not uniquely determined from SVD of the matrix GP. The recovery of rotation matrices is performed by constructing for every $i = 1, ..., K$ a 3×3 matrix

$$A_i = \begin{pmatrix} | & | & | \\ A_i^1 & A_i^2 & A_i^3 \\ | & | & | \end{pmatrix}$$

whose columns are given by

$$A_i^1 = \begin{pmatrix} u_{2i-1}^1 \\ u_{2i-1}^2 \\ u_{2i-1}^3 \end{pmatrix},\ A_i^2 = \begin{pmatrix} u_{2i}^1 \\ u_{2i}^2 \\ u_{2i}^3 \end{pmatrix},\ A_i^3 = \text{cross}\left(A_i^1, A_i^2\right).$$

In practice, due to erroneous common lines, the vectors A_i^1 and A_i^2 are approximately two orthonormal vectors with unit norm. Define 3×2 matrices constructed from A_i and R_i as

$$A_i^{[1,2]} = \begin{pmatrix} | & | \\ A_i^1 & A_i^2 \\ | & | \end{pmatrix} \text{ and } R_i^{[1,2]} = \begin{pmatrix} | & | \\ R_i^1 & R_i^2 \\ | & | \end{pmatrix}.$$

We estimate the matrix $R_i^{[1,2]}$ as the closest matrix to $A_i^{[1,2]}$ on the Stiefel manifold $V_2(\mathbb{R}^3)$ in the Frobenius matrix norm. This is done via the well-known procedure [1]. $R_i^{[1,2]} = U_i V_i^T$, where $A_i^{[1,2]} = U_i \Sigma_i V_i^T$ is the singular value decomposition of $A_i^{[1,2]}$. We note that except for the orthogonality constraint (4.7), the semidefinite program (4.6)–(4.8) is identical to the Goemans–Williamson SDP for finding the maximum cut in a weighted graph [11], where the SDR and the randomized rounding procedure [31, 15] for maximum cut problem is proved to have a 0.87 performance guaranty. From the complexity point of view, SDP can be solved in polynomial time to any given precision. The idea of using SDP for determining image orientations in cryo-EM was originally proposed in [29].

5. The Spectral Norm Constraint. However in our numerical experiments (section 8), we observed that when the detected common-lines have a lot of "out-liers", the viewing directions of images estimated by solving either (4.6)-(4.8) or (4.10) are highly clustered (Figure 5.1). In other words, instead of being a global assignment of rotations in $\mathbf{SO}(3)$ for the images orientation, the solution to (4.6)-(4.8) or (4.10) tends to be a global assignment of 2D orthogonal transformation in $\mathbf{O}(2)$ which is a combination of in-plane rotations and reflections. In order to prevent such tendency, we add another constraint on the spectral norm of the Gram matrix G to the optimization problem (4.6)-(4.8) or (4.10).

$$G \preccurlyeq \alpha K \cdot I \tag{5.1}$$

or equivalently

$$\|G\|_2 \leq \alpha K, \tag{5.2}$$

where $\|G\|_2$ is the spectral norm of the matrix G, and $\frac{2}{3} \leq \alpha < 1$ controls the spread of the viewing directions. If the true image orientations are uniformly sampled from the rotation group $\mathbf{SO}(\mathbf{3})$, then by the law of large number and the symmetry of the distribution of orientations, we can prove the spectrum norm of the true Gram matrix

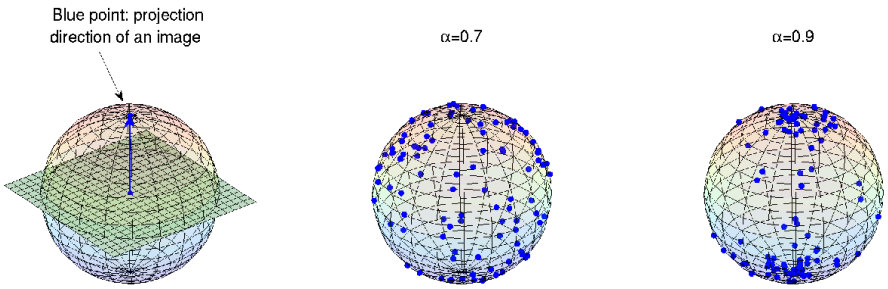


Fig. 5.1: The dependency of the spectral norm of G (denoted as αK here) on the distribution of orientations of the images. Here $K = 100$. The larger α is, the more clustered the orientations are.

G_{true} is approximately $\frac{2}{3}K$. If the true viewing directions are highly clustered, then the spectral norm of the true Gram matrix G_{true} is approximately K. For a known distribution of orientations, we can compute the spectrum norm of the true Gram matrix G_{true} accordingly, which can be verified to be a number between $\frac{2}{3}$ and 1. To prevent a solution of clustered viewing angles, we fix α to some number satisfying $\frac{2}{3} \leq \alpha < 1$.

6. The Alternating Direction Method (ADM) for the problems with the Spectral Norm Constraint. Here we generalize the application of ADM to SDP problems in [40] to the application to the problems raised in the previous sections. ADM is a multiple-splitting algorithm that minimizes the dual augmented Lagrangian function in an alternating fashion such that in each step it firstly minimizes the Lagrange multipliers, then the dual slack variables, and finally the primal variables. In addition, in the minimization over a certain variable, the other variables are kept fixed.

6.1. ADM for the relaxed weighted LS problem with the Spectral Norm Constraint. The weighted LS problem after SDR (4.6)-(4.8) can be efficiently solved using SDPLR [3]. However, SDPLR is not suitable for the problem after the spectral norm constraint on G (5.2) is added to (4.6)-(4.8). This is because the constraint (5.2) can be written as $\alpha KI - G \succcurlyeq 0$ where $\alpha KI - G$ does not have a low rank structure. Moreover, SDP solvers using polynomial-time primal-dual interior point methods are designed for small to medium sized problems. Therefore they are not suitable for our problem. Here we devise a ADM for this problem which takes advantage of the low-rank property of G. After the spectral norm constraint (5.2) is added, the problem becomes

$$\min_{G \succcurlyeq 0} \quad -\langle C, G \rangle \tag{6.1}$$

$$\text{s.t. } \mathcal{A}(G) = \mathbf{b} \tag{6.2}$$

$$\|G\|_2 \leq \alpha K \tag{6.3}$$

where

$$\mathcal{A}(G) = \begin{pmatrix} G_{ii}^{11} \\ G_{ii}^{22} \\ \frac{\sqrt{2}}{2} G_{ii}^{12} + \frac{\sqrt{2}}{2} G_{ii}^{21} \end{pmatrix}_{i=1,2,\ldots,N}, \ \mathbf{b} = \begin{pmatrix} b_i^1 \\ b_i^2 \\ b_i^3 \end{pmatrix}_{i=1,2,\ldots,N}, \tag{6.4}$$

$$b_i^1 = b_i^2 = 1,\ b_i^3 = 0 \text{ for all } i,$$

G_{ij}^{pq} denotes the (p,q) th element in the 2×2 sub-block G_{ij}, $C = W \circ S$ is a symmetric matrix and the product $\langle C, G\rangle = \text{trace}\,(CG)$. Following the equality $\langle \mathcal{A}(G), \mathbf{y}\rangle = \langle G, \mathcal{A}^*(\mathbf{y})\rangle$ for arbitrary $\mathbf{y} = \left(\begin{array}{c} y_i^1 \\ y_i^2 \\ y_i^3 \end{array} \right)_{i=1,2,\ldots,N}$, the adjoint of the operator $\mathcal{A}$ is defined as

$$\mathcal{A}^*(\mathbf{y}) = Y = \left(\begin{array}{cc} Y^{11} & Y^{12} \\ Y^{21} & Y^{22} \end{array} \right),$$

where for $i = 1, 2, \ldots, N$

$$Y_{ii}^{11} = y_i^1,\ Y_{ii}^{22} = y_i^2, \text{ and } Y_{ii}^{12} = Y_{ii}^{21} = y_i^3/\sqrt{2}.$$

It can be verified that $\mathcal{A}\mathcal{A}^* = I$. The dual problem of problem (6.1)-(6.3) is

$$\max_{\mathbf{y}, X\succcurlyeq 0} \min_{\|G\|_2 \le \alpha K} -\langle C, G\rangle - \langle \mathbf{y}, \mathcal{A}(G) - \mathbf{b}\rangle - \langle G, X\rangle . \tag{6.5}$$

By rearranging terms in (6.5), we obtain

$$\max_{\mathbf{y}, X\succcurlyeq 0} \min_{\|G\|_2 \le \alpha K} -\langle C + X + \mathcal{A}^*(\mathbf{y}), G\rangle + \mathbf{y}^T\mathbf{b}. \tag{6.6}$$

Using the fact that the dual norm of the spectral norm is nuclear norm (Proposition 2.1 in [25]), we can obtain from (6.6) the dual problem

$$\max_{\mathbf{y}, X\succcurlyeq 0} \mathbf{y}^T\mathbf{b} - \alpha K \|C + X + \mathcal{A}^*(\mathbf{y})\|_* , \tag{6.7}$$

where $\|\cdot\|_*$ denotes the nuclear norm. Introducing a variable $Z = C + X + \mathcal{A}^*(\mathbf{y})$, we obtain from (6.7) that

$$\min_{\mathbf{y}, X\succcurlyeq 0} \quad -\mathbf{y}^T\mathbf{b} + \alpha K \|Z\|_* \tag{6.8}$$

$$\text{s.t.} Z = C + X + \mathcal{A}^*(\mathbf{y}) . \tag{6.9}$$

Since Z is a symmetric matrix, $\|Z\|_*$ is the summation of the absolute value of the eigenvalues of Z. The augmented Lagrangian function of (6.8)-(6.9) is defined as

$$\begin{aligned} \mathcal{L}(\mathbf{y}, Z, X, G) = &-\mathbf{y}^T\mathbf{b} + \alpha K \|Z\|_* + \langle G, C + X + \mathcal{A}^*(\mathbf{y}) - Z\rangle \\ &+ \frac{\mu}{2}\|C + X + \mathcal{A}^*(\mathbf{y}) - Z\|_F^2 , \end{aligned} \tag{6.10}$$

where $\mu > 0$ is a penalty parameter. Using the augmented Lagrangian function (6.10), we devise an ADM that minimizes (6.10) with respect to $\mathbf{y}$, Z, X, and G in an alternating fashion, that is, given some initial guess, in each iteration the following three subproblems are solved sequentially:

$$\mathbf{y}^{k+1} = \arg\min_{\mathbf{y}} \mathcal{L}\left(\mathbf{y}, Z^k, X^k, G^k\right), \tag{6.11}$$

$$Z^{k+1} = \arg\min_{Z} \mathcal{L}\left(\mathbf{y}^{k+1}, Z, X^k, G^k\right), \tag{6.12}$$

$$X^{k+1} = \arg\min_{X\succcurlyeq 0} \mathcal{L}\left(\mathbf{y}^{k+1}, Z^{k+1}, X, G^k\right), \tag{6.13}$$

and the Lagrange multiplier G is updated by

$$G^{k+1} = G^k + \gamma\mu\left(C + X^{k+1} + \mathcal{A}^*\left(\mathbf{y}^{k+1}\right) - Z^{k+1}\right), \tag{6.14}$$

where $\gamma \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ is an appropriately chosen step length.

To solve the subproblem (6.11), we use the first order optimality condition

$$\nabla_{\mathbf{y}}\mathcal{L}\left(\mathbf{y}, Z^k, X^k, G^k\right) = 0$$

and the fact that $\mathcal{A}\mathcal{A}^* = I$, and we obtain

$$\mathbf{y}^{k+1} = -\mathcal{A}\left(C + X^k - Z^k\right) - \frac{1}{\mu}\left(\mathcal{A}\left(G\right) - \mathbf{b}\right).$$

By rearranging the terms of $\mathcal{L}\left(\mathbf{y}^{k+1}, Z, X^k, G^k\right)$, it can be verified that the subproblem (6.12) is equivalent to

$$\min_Z \frac{\alpha K}{\mu}\left\|Z\right\|_* + \frac{1}{2}\left\|Z - B^k\right\|_F^2,$$

where $B^k = C + X^k + \mathcal{A}^*\left(\mathbf{y}^{k+1}\right) + \frac{1}{\mu}G^k$. Let $B^k = U\Lambda U^T$ be the spectral decomposition of the matrix B^k, where $\Lambda = \operatorname{diag}\left(\boldsymbol{\lambda}\right) = \operatorname{diag}\left(\lambda_1, \ldots, \lambda_{2K}\right)$. Then $Z^{k+1} = U\operatorname{diag}\left(\hat{\mathbf{z}}\right)U^T$, where $\hat{\mathbf{z}}$ is the optimal solution of the problem

$$\min_{\mathbf{z}} \frac{\alpha K}{\mu}\left\|\mathbf{z}\right\|_1 + \frac{1}{2}\left\|\mathbf{z} - \boldsymbol{\lambda}\right\|_2^2, \tag{6.15}$$

It can be shown that the unique solution of (6.15) admits a closed form called the soft-thresholding operator, following a terminology introduced by Donoho and Johnstone [5]; it can be written as

$$\hat{z}_i = \begin{cases} 0, & \text{if } |\lambda_i| \leq \alpha K/\mu \\ (1 - \frac{\alpha}{\mu}/|\lambda_i|)\lambda_i, & \text{otherwise.} \end{cases}$$

The problem (6.13) can be shown to be equivalent to

$$\min_X \left\|X - H^k\right\|_F^2, \text{ s.t. } X \succcurlyeq 0,$$

where $H^k = Z^{k+1} - C - \mathcal{A}^*\left(\mathbf{y}^{k+1}\right) - \frac{1}{\mu}G^k$. The solution $X^{k+1} = V_+\Sigma_+V_+^T$ is the Euclidean projection of H^k onto the semidefinite cone (section 8.1.1 in [2]), where

$$V\Sigma V^T = \begin{pmatrix} V_+ & V_- \end{pmatrix}\begin{pmatrix} \Sigma_+ & 0 \\ 0 & \Sigma_- \end{pmatrix}\begin{pmatrix} V_+^T \\ V_-^T \end{pmatrix}$$

is the spectral decomposition of the matrix H^k, and Σ_+ and Σ_- are the positive and negative eigenvalues of H^k.

It follows from the updating equation (6.14) that

$$\begin{aligned} G^{k+1} &= (1-\gamma)G^k + \gamma\mu\left(C + X^{k+1} + \mathcal{A}^*\left(\mathbf{y}^{k+1}\right) - Z^{k+1} + \frac{1}{\mu}G^k\right) \\ &= (1-\gamma)G^k + \gamma\mu\left(X^{k+1} - H^k\right). \end{aligned}$$

6.2. ADM for the relaxed LUD problem with the Spectral Norm Constraint. Consider the LUD problem after SDR:

$$\min_{G\succcurlyeq 0}\sum_{i<j}\left\|\vec{c}_{ij}^{T}-G_{ij}\vec{c}_{ji}^{T}\right\| \text{ s.t. } \mathcal{A}(G)=\mathbf{b}, \tag{6.16}$$

where G, $\mathcal{A}$ and $\mathbf{b}$ are defined in (4.3) and (6.4) respectively. The ADM devised to solve (6.16) is similar to and simpler than the ADM devised to solve the one with the spectral norm constraint. We focus on the more difficult problem with the spectral norm constraint. Introducing $\mathbf{x}_{ij}=\vec{c}_{ij}^{T}-G_{ij}\vec{c}_{ji}^{T}$ and adding the spectral norm constraint $\|G\|_2\le\alpha K$, we obtain

$$\min_{\mathbf{x}_{ij},G\succcurlyeq 0}\sum_{i<j}\|\mathbf{x}_{ij}\| \text{ s.t. } \mathcal{A}(G)=\mathbf{b},\ \mathbf{x}_{ij}=\vec{c}_{ij}^{T}-G_{ij}\vec{c}_{ji}^{T},\ \|G\|_2\le\alpha K. \tag{6.17}$$

The dual problem of problem (6.17) is

$$\max_{\boldsymbol{\theta}_{ij},\mathbf{y},X\succcurlyeq 0}\ \min_{\mathbf{x}_{ij},\|G\|_2\le\alpha K}\sum_{i<j}\left(\|\mathbf{x}_{ij}\|-\left\langle\boldsymbol{\theta}_{ij},\mathbf{x}_{ij}-\vec{c}_{ij}^{T}+G_{ij}\vec{c}_{ji}^{T}\right\rangle\right)-\langle\mathbf{y},\mathcal{A}(G)-\mathbf{b}\rangle-\langle G,X\rangle. \tag{6.18}$$

By rearranging terms in (6.18) , we obtain

$$\begin{aligned}\max_{\boldsymbol{\theta}_{ij},\mathbf{y},X\succcurlyeq 0}\ \min_{\mathbf{x}_{ij},\|G\|_2\le\alpha K} & -\langle Q(\boldsymbol{\theta})+X+\mathcal{A}^{*}(\mathbf{y}),G\rangle+\mathbf{y}^{T}\mathbf{b}\\ & +\sum_{i<j}\left(\|\mathbf{x}_{ij}\|-\langle\boldsymbol{\theta}_{ij},\mathbf{x}_{ij}\rangle+\left\langle\boldsymbol{\theta}_{ij},\vec{c}_{ij}^{T}\right\rangle\right),\end{aligned} \tag{6.19}$$

where $\boldsymbol{\theta}=(\boldsymbol{\theta}_{ij})_{i,j=1,\ldots,K}$, $\boldsymbol{\theta}_{ij}=\left(\theta_{ij}^{1},\theta_{ij}^{2}\right)^{T}$, $\vec{c}_{ij}=\left(c_{ij}^{1},c_{ij}^{2}\right)$,

$$Q(\boldsymbol{\theta})=\frac{1}{2}\begin{pmatrix}Q^{11}(\boldsymbol{\theta}) & Q^{12}(\boldsymbol{\theta})\\ Q^{21}(\boldsymbol{\theta}) & Q^{22}(\boldsymbol{\theta})\end{pmatrix} \text{ and } Q^{pq}(\boldsymbol{\theta})=\begin{pmatrix}0 & \theta_{12}^{p}c_{21}^{q} & \cdots & \theta_{1K}^{p}c_{K1}^{q}\\ c_{21}^{q}\theta_{12}^{p} & 0 & \cdots & \theta_{2K}^{p}c_{K2}^{q}\\ \vdots & \vdots & \ddots & \vdots\\ c_{K1}^{q}\theta_{1K}^{p} & c_{K2}^{q}\theta_{2K}^{p} & \cdots & 0\end{pmatrix}$$

for p, $q=1,2$. It is easy to verify that for $1\le i<j\le K$

$$\min_{\mathbf{x}_{ij}}\left(\|\mathbf{x}_{ij}\|-\langle\boldsymbol{\theta}_{ij},\mathbf{x}_{ij}\rangle\right)=\begin{cases}0 & \text{if } \|\boldsymbol{\theta}_{ij}\|\le 1\\ -\infty & \text{otherwise.}\end{cases} \tag{6.20}$$

In fact, (6.20) is obtained using the inequality

$$\begin{aligned}\|\mathbf{x}_{ij}\|-\langle\boldsymbol{\theta}_{ij},\mathbf{x}_{ij}\rangle &= \|\mathbf{x}_{ij}\|-\|\boldsymbol{\theta}_{ij}\|\|\boldsymbol{x}_{ij}\|\left\langle\boldsymbol{\theta}_{ij}/\|\boldsymbol{\theta}_{ij}\|,\mathbf{x}_{ij}/\|\boldsymbol{x}_{ij}\|\right\rangle\\ &\ge \|\mathbf{x}_{ij}\|-\|\boldsymbol{\theta}_{ij}\|\|\boldsymbol{x}_{ij}\|=(1-\|\boldsymbol{\theta}_{ij}\|)\|\boldsymbol{x}_{ij}\|,\end{aligned} \tag{6.21}$$

and the inequality in (6.21) holds when $\boldsymbol{\theta_{ij}}$ and $\mathbf{x}_{ij}$ have the same direction. Using the fact that the dual norm of the spectral norm is nuclear norm and the fact in (6.20), we can obtain from (6.19) the dual problem

$$\min_{\boldsymbol{\theta}_{ij},\mathbf{y},X\succcurlyeq 0}\ -\mathbf{y}^{T}\mathbf{b}-\sum_{i<j}\left\langle\boldsymbol{\theta}_{ij},\vec{c}_{ij}^{T}\right\rangle+\alpha K\|Z\|_{*} \tag{6.22}$$

$$\text{s.t.}\qquad Z=Q(\boldsymbol{\theta})+X+\mathcal{A}^{*}(\mathbf{y}),\qquad \text{and } \|\boldsymbol{\theta}_{ij}\|\le 1. \tag{6.23}$$

The augmented Lagrangian function of problem (6.22)-(6.23) is defined as

$$\mathcal{L}(\mathbf{y},\boldsymbol{\theta},Z,X,G) = -\mathbf{y}^T\mathbf{b} + \alpha K\|Z\|_* - \sum_{i<j}\langle \boldsymbol{\theta}_{ij}, \vec{c}_{ij}^T\rangle + \langle G, Q(\boldsymbol{\theta}) + X + \mathcal{A}^*(\mathbf{y}) - Z\rangle + \frac{\mu}{2}\|Q(\boldsymbol{\theta}) + X + \mathcal{A}^*(\mathbf{y}) - Z\|_F^2, \tag{6.24}$$

for $\|\boldsymbol{\theta}_{ij}\| \leq 1$, where $\mu > 0$ is a penalty parameter. Similar to section 6.1, using the augmented Lagrangian function (6.24), ADM is used to minimize (6.24) with respect to $\mathbf{y}$, $\boldsymbol{\theta}$, Z, X, and G alternatively, that is, given some initial guess, in each iteration the following four subproblems are solved sequentially:

$$\mathbf{y}^{k+1} = \arg\min_{\mathbf{y}} \mathcal{L}\left(\mathbf{y}, \boldsymbol{\theta}^k, Z^k, X^k, G^k\right), \tag{6.25}$$

$$\boldsymbol{\theta}_{ij}^{k+1} = \arg\min_{\|\boldsymbol{\theta}_{ij}\|\leq \mathbf{1}} \mathcal{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}, Z^k, X^k, G^k\right), \tag{6.26}$$

$$Z^{k+1} = \arg\min_{Z} \mathcal{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}^{k+1}, Z, X^k, G^k\right), \tag{6.27}$$

$$X^{k+1} = \arg\min_{X\succcurlyeq 0} \mathcal{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}^{k+1}, Z^{k+1}, X, G^k\right), \tag{6.28}$$

and the Lagrange multiplier G is updated by

$$G^{k+1} = G^k + \gamma\mu\left(Q\left(\boldsymbol{\theta}^{k+1}\right) + X^{k+1} + \mathcal{A}^*\left(\mathbf{y}^{k+1}\right) - Z^{k+1}\right), \tag{6.29}$$

where $\gamma \in \left(0, \frac{1+\sqrt{5}}{2}\right)$ is an appropratly chosen step length. The methods to solve subproblems (6.25), (6.27) and (6.28) are similar to those used in (6.11), (6.12) and (6.13. To solve subproblem (6.26), we rearrange the terms of $\mathcal{L}\left(\mathbf{y}^{k+1}, \boldsymbol{\theta}, Z^k, X^k, G^k\right)$ and obtain an eqivalent problem

$$\min_{\boldsymbol{\theta}_{ij}} -\langle \boldsymbol{\theta}_{ij}, \vec{c}_{ij}^T\rangle + \frac{\mu}{2}\|\boldsymbol{\theta}_{ij}\vec{c}_{ji} + \Phi_{ij}\|_F^2, \text{ s.t. } \|\boldsymbol{\theta}_{ij}\| \leq 1,$$

where $\Phi = X^k + \mathcal{A}^*\left(\mathbf{y}^{k+1}\right) - Z^k + \frac{1}{\mu}G^k$, $\Phi = \begin{pmatrix} \Phi^{11} & \Phi^{12} \\ \Phi^{21} & \Phi^{22} \end{pmatrix}$ and $\Phi_{ij} = \begin{pmatrix} \Phi_{ij}^{11} & \Phi_{ij}^{12} \\ \Phi_{ij}^{21} & \Phi_{ij}^{22} \end{pmatrix}$. Problem (6.29) is further simplified as

$$\min_{\boldsymbol{\theta}_{ij}} \langle \boldsymbol{\theta}_{ij}, \mu\Phi_{ij}\vec{c}_{ji}^T - \vec{c}_{ij}^T\rangle + \frac{\mu}{2}\|\boldsymbol{\theta}_{ij}\|^2, \text{ s.t. } \|\boldsymbol{\theta}_{ij}\| \leq 1,$$

whose solution is

$$\boldsymbol{\theta}_{ij} = \begin{cases} \frac{1}{\mu}\vec{c}_{ij}^T - \Phi_{ij}\vec{c}_{ij}^T & \text{if } \left\|\frac{1}{\mu}\vec{c}_{ij}^T - \Phi_{ij}\vec{c}_{ij}^T\right\| \leq 1, \\ \frac{\vec{c}_{ij}^T - \mu\Phi_{ij}\vec{c}_{ij}^T}{\|\vec{c}_{ij}^T - \mu\Phi_{ij}\vec{c}_{ij}^T\|} & \text{otherwise.} \end{cases}$$

The convergence analysis and the practical issues related to how to take advantage of low-rank assumption of G in the eigenvalue decomposition performed at each iteration, strategies for adjusting the penalty parameter μ, the use of a step size γ for updating the primal variable X and termination rules using the in-feasibility measures are discussed in details in [40].

7. The Iterative Reweighted Least Squares (IRLS) Procedure. The LUD problem (3.7) can also be solved via semidefinite relaxation as

$$\min_{G\in\mathbb{R}^{2K\times 2K}} F(G) = \sum_{i,j=1,2,\ldots,K} \sqrt{2 - 2\sum_{p,q=1,2} G_{ij}^{pq} S_{ij}^{pq}} \tag{7.1}$$

$$\text{s.t.} \qquad G_{ii} = I_2,\ i = 1, 2, \cdots, K, \tag{7.2}$$

$$G \succcurlyeq 0, \tag{7.3}$$

$$\|G\|_2 \le \alpha K \text{ (optional)}, \tag{7.4}$$

where α is a fixed number between $\frac{2}{3}$ and 1, and the spectral norm constraint on G (7.4) is added when the solution to the problem (7.1)-(7.3) is a set of highly clustered rotations. Notice that this relaxed problem is not convex since the objective function (7.1) is concave. Thus a good initial guess of G, for example, an estimate of G using LS, is used to obtain an approximate solution of (7.1)-(7.3) (possibly with (7.4)) by the IRLS procedure [4, 14] described in algorithm 1. Before the rounding procedure, the

Algorithm 1 (**the IRLS procedure**) Solve optimization problem (7.1)-(7.3) (with the spectral norm constraint on G (7.4) if the input parameter α satisfies $\frac{2}{3} \le \alpha < 1$), and then recover the orientations by rounding.

Require: a $2K \times 2K$ common-line matrix S, a regularization parameter ϵ, a parameter α and the total number of iterations N

$w_{ij}^0 = 1\ \forall i, j = 1, \cdots, K$;

$G^0 = 0$;

for $k = 1 \to N$, step size $= 1$ **do**

update W by setting $w_{ij} = w_{ij}^{k-1}$;

if $\frac{2}{3} \le \alpha < 1$, obtain G^k by solving the problem (6.1)-(6.3) using ADM; otherwise, obtain G^k by solving (4.6)–(4.8) using SDPLR (with initial guess G^{k-1});

$r_{ij}^k = \sqrt{2 - 2\sum_{p,q=1,2} G_{ij}^{pq} S_{ij}^{pq} + \epsilon^2}$;

$w_{ij}^k = 1/r_{ij}^k$;

the residual $r^k = \sum_{i,j=1}^K r_{ij}^n$;

end for

obtain estimated orientations $\hat{R}_1, \ldots, \hat{R}_K$ from G^N using the randomized rounding procedure in section 4.4.

IRLS procedure finds an approximate solution to the optimization problem (7.1)-(7.3) (possibly with (7.4)) by solving its smoothing version

$$\min_{G\in\mathbb{R}^{2K\times 2K}} F(G,\epsilon) = \sum_{i,j=1,2,\ldots,K} \sqrt{2 - 2\sum_{p,q=1,2} G_{ij}^{pq} S_{ij}^{pq} + \epsilon^2} \tag{7.5}$$

$$\text{s.t.} \qquad G_{ii} = I_2,\ i = 1, 2, \cdots, K, \tag{7.6}$$

$$G \succcurlyeq 0, \tag{7.7}$$

$$\|G\|_2 \le \alpha K \text{ (optional)}. \tag{7.8}$$

where $\epsilon > 0$ is a small number. In each iteration, the IRLS procedure solves a problem

$$G^{k+1} = \arg\min_{G\succcurlyeq 0} \sum_{i\neq j} w_{ij}^k \left(2 - 2\left\langle G_{ij}, S_{ij}\right\rangle + \epsilon^2\right) \text{ s.t. } \mathcal{A}(G) = \mathbf{b}, (\text{optional: } \|G\|_2 \le \alpha K) \tag{7.9}$$

on the $(k+1)$th iteration, where $w_{ij}^0 = 1$, and

$$w_{ij}^k = 1/\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2},\ \forall k > 0.$$

In other words, we emphasize the detected common-lines that are explained well by the current estimated Gram matrix G^k. The presence of the regularization ϵ ensures that no single detected common-line can gain undue influence.

$$G^{k+1} = \arg\max_{G \succcurlyeq 0} \left\langle W^k \circ S, G\right\rangle \text{ s.t. } \mathcal{A}(G) = \mathbf{b} \text{ (optional: } \|G\|_2 \leq \alpha K). \tag{7.10}$$

We repeat the process until the residual sequence $\{r^k\}$ has converged. We prove the following theorem.

THEOREM 7.1. *Let $\left\{G^k\right\}$ be the sequence generated by the IRLS procedure, then*

$$F(G^{k+1}, \epsilon) \leq F(G^k, \epsilon). \tag{7.11}$$

Proof. Here we prove the theorem without the spectral norm constraint on G. The arguments can be generalized to the IRLS procedure with the spectral norm constraint. Since G^k is the solution of (7.10), there exists $\mathbf{y}^k \in \mathbb{R}^{2n}$ and $X^k \in \mathbb{R}^{2n\times 2n}$ such that

$$-\mathcal{A}^*(\mathbf{y}^k) + X^k + W^{k-1} \circ S = 0,\ \mathcal{A}(G^k) - \mathbf{b} = 0, \tag{7.12}$$

$$G^k \succcurlyeq 0,\ X^k \succcurlyeq 0,\ \left\langle G^k, X^k\right\rangle = 0. \tag{7.13}$$

Hence we have

$$\begin{aligned}
0 &= -(\mathbf{y}^k)^T(\mathcal{A}(G^k) - \mathbf{b}) + (\mathbf{y}^{k+1})^T(\mathcal{A}(G^{k+1}) - \mathbf{b}) \\
&= (\mathbf{y}^{k+1} - \mathbf{y}^k)^T(\mathcal{A}(G^k) - \mathbf{b}) + \left\langle \mathcal{A}^*(\mathbf{y}^{k+1}), G^{k+1} - G^k\right\rangle \\
&= \left\langle X^{k+1} + W^k \circ S, G^{k+1} - G^k\right\rangle \\
&\leq \left\langle W^k \circ S, G^{k+1} - G^k\right\rangle \qquad (7.14) \\
&= \frac{1}{2}\sum_{i\neq j}\left(-\beta_{ij}^k\left(2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2\right) + \beta_{ij}^k\left(2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2\right)\right) \\
&= \frac{1}{2}\sum_{i\neq j}\left(-\frac{2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2}{\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2}} + \sqrt{2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2}\right), \qquad (7.15)
\end{aligned}$$

where the third equality uses (7.12), and the inequality (7.14) uses (7.13). From (7.15) we obtain

$$\begin{aligned}
F(G^k, \epsilon)^2 &= \left(\sum_{i\neq j}\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2}\right)^2 \\
&\geq \left(\sum_{i\neq j}\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2}\right)\left(\sum_{i\neq j}\frac{2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2}{\sqrt{2 - 2\left\langle G_{ij}^k, S_{ij}\right\rangle + \epsilon^2}}\right) \\
&\geq \left(\sum_{i\neq j}\sqrt{2 - 2\left\langle G_{ij}^{k+1}, S_{ij}\right\rangle + \epsilon^2}\right)^2 = F(G^{k+1}, \epsilon)^2, \qquad (7.16)
\end{aligned}$$

where the last inequality uses Cauchy-Schwarz inequality and the equality holds if and only if

$$\frac{\sqrt{2-2\left\langle G_{ij}^{k+1},S_{ij}\right\rangle+\epsilon^2}}{\sqrt{2-2\left\langle G_{ij}^{k},S_{ij}\right\rangle+\epsilon^2}}=c \text{ for all } i\neq j, \tag{7.17}$$

where c is a constant. Thus (7.11) is confirmed. □

In addition, using Hölder's inequality, the analysis can be generalized to the reweighted approach to solve

$$\min_{G\succcurlyeq 0}\sum_{i\neq j}\left(2-2\left\langle G_{ij},S_{ij}\right\rangle\right)^{\frac{p}{2}} \text{ s.t. } \mathcal{A}(G)=\mathbf{b}, (\text{optional: } \|G\|_2\leq \alpha K) \tag{7.18}$$

where $0<p<1$. The problem (7.18) is a SDR of the problem

$$\min_{R_1,\ldots,R_K\in\mathbf{SO}(3)}\sum_{i\neq j}\left\|R_i\left(\vec{c}_{ij},0\right)^T-R_j\left(\vec{c}_{ji},0\right)^T\right\|^p. \tag{7.19}$$

The smaller p is, the more penalty the outliers in the detected common-lines receive.

8. Numerical results. All numerical experiments were performed on a machine with 2 Intel(R) Xeon(R) CPUs X5570, each with 4 cores, running at 2.93 GHz. We simulated 100 centered images of size 129×129 pixels of the 50S ribosomal subunit with SNR=1/32 and 1/64 respectively (Figure 8.1). The noise added to the images is white Gaussian. The pixel size is 2.4Å. The orientations of the images are sampled from the uniform distribution over $\mathbf{SO}(3)$. The polar Fourier transform have radial resolution $n_r=100$ and angular resolution $n_\theta=360$. Common-line pairs that are detected with an error smaller than 10° are considered to be correct. The common-line detection rates are 44% and 23% for images with SNR=1/32 and SNR=1/64, respectively (Figure 3.1).

We define the mean squared error (MSE) of the estimated rotation matrices $\hat{R}_1,\ldots,\hat{R}_K$ as

$$\text{MSE}=\frac{1}{K}\sum_{i=1}^{K}\left\|R_i-\hat{O}\hat{R}_i\right\|^2, \tag{8.1}$$

where $\hat{O}$ is the optimal solution to the registration problem between the two sets of rotations $\{R_1,\ldots,R_K\}$ and $\left\{\hat{R}_1,\ldots,\hat{R}_K\right\}$ in the sense of minimizing the MSE. As shown in [29], there is a simple procedure to obtain both $\hat{O}$ and the MSE from the singular value decomposition of the matrix $\frac{1}{K}\sum_{i=1}^{K}\hat{R}_iR_i^T$.

When SNR=1/32, the common-line detection rate is relatively high (44%), the spectral constraint on G is not needed. The LUD approach using ADM and IRLS outweighs the LS approach in terms of accuracy (Figure 8.2 and Table 8.1a). When SNR=1/64, the common-line detection rate is relatively small (23%), the spectral constraint on G improves the accuracy of estimations for all algorithms (Table 8.1b). The LUD approach is better than the LS approach (Figure 8.2 and Table 8.1b). In addition, the LUD approach using ADM is less sensitive to the choice of the parameter α than the one using IRLS (Table 8.1b). We also observed that increasing the number of images helps little in improving the MSEs.

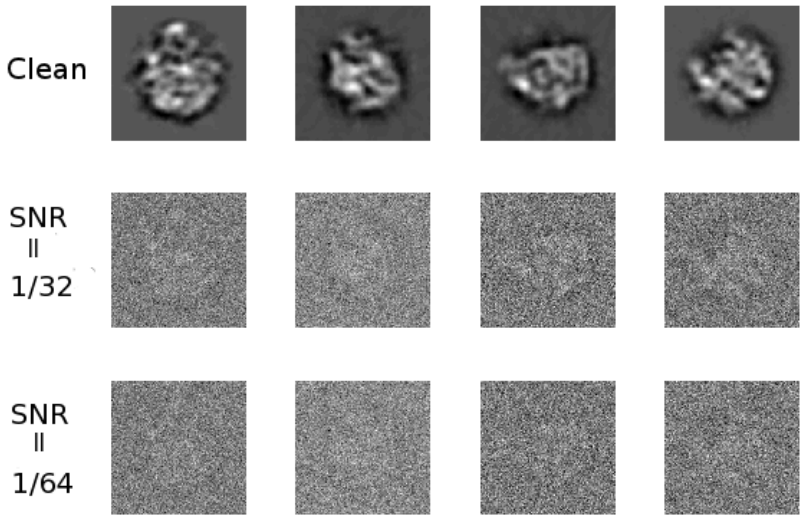


Fig. 8.1: The top row shows four clean images of size 129×129 pixels generated from a 50S ribosomal subunit volume with different orientations. The middle and bottom rows show four noisy images corresponding to those in the top row with SNR=1/32 and 1/64, respectively.

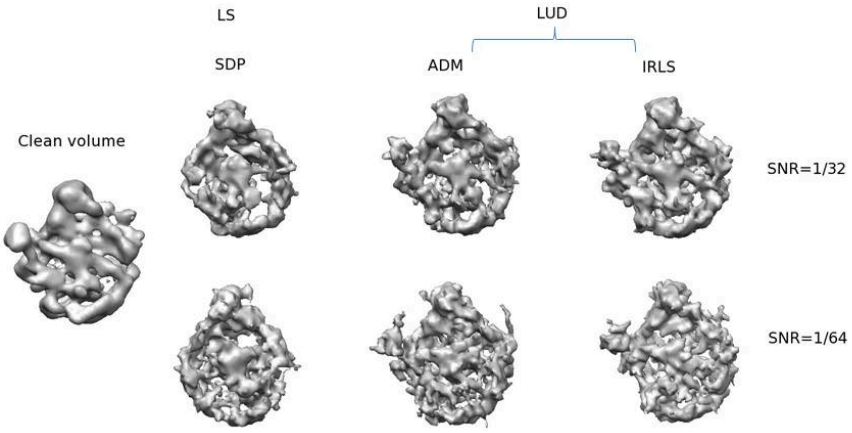


Fig. 8.2: The clean volume (leftmost) and the reconstructed volumes. For images with $SNR = 1/64$, we used the spectral constraint on G and set $\alpha = 2/3$ for all algorithms.

	Method	MSE	Time
LS	SDP	0.36	2s
LUD	ADM	0.14	19s
	IRLS	**0.06**	23s

(a) $SNR = 1/32$

	Method	α	MSE	Time
LS	SDP	NA	1.09	2s
	ADM	2/3	0.74	12s
		3/4	0.75	13s
		4/5	0.82	17s
LUD	ADM	NA	2.01	11s
		2/3	0.69	42s
		3/4	0.67	42s
		4/5	0.66	45s
	IRLS	NA	2.12	48s
		2/3	**0.56**	432s
		3/4	0.66	267s
		4/5	0.74	295s

(b) $SNR = 1/64$

Table 8.1: The MSEs (8.1) of the estimated rotations and the cost time using different algorithms.

9. Discussion. To estimate image orientations, we introduced a robust self consistency error and used ADM or the IRLS procedure to solve the associated LUD problem after SDR. Numerical experiments demonstrate that the solution is less sensitive to outliers in the detected common-lines than the LS method approach. In addition,when the common-line detection rate is low, the spectral norm constraint on the Gram matrix G introduced helps tighten the semidefinite relaxation, and thus improves the accuracy of the result. We note that it is also possible to consider other self consistency errors involving the unsquared deviations raised to some power p (e.g., the cases $p = 1, 2$ correspond to LUD and LS, respectively). We observed that the accuracy of the estimated orientations can be improved by using $p < 1$ provided that the initial guess is "good enough". The LUD approach and the spectral norm constraint on G can be generalized to the synchronization approach to estimate the images' orientations in [27]. It is speculated that under some condition, the exact recovery of the orientations can be achieved using the LUD approach as discussed in our another paper on synchronization over rotation groups [39].

10. Acknowledgements. This work was partially supported by Award Number DMS-0914892 from the NSF, by Award Number FA9550-09-1-0551 from AFOSR, by Award Number R01GM090200 from the NIGMS and by the Alfred P. Sloan Foundation.

REFERENCES

[1] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(5):698–700, May 1987.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[3] S. Burer and R. D. C. Monteiro. A Nonlinear Programming Algorithm for Solving Semidefinite Programs via Low-rank Factorization. *Mathematical Programming (series B)*, 95:2003, 2001.

[4] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, 2010.

[5] D. L. Donoho and I. M. Johnstone. Adapting to Unknown Smoothness via Wavelet Shrinkage. *Journal of the American Statistical Association*, 90(432):1200+, December 1995.

[6] A. Dutt and V. Rokhlin. Fast Fourier Transforms for Nonequispaced Data. *SIAM Journal on Scientific Computing*, 14(6):1368–1393, 1993.

[7] N. A. Farrow and F. P. Ottensmeyer. A posteriori determination of relative projection directions of arbitrarily oriented macromolecules. *J. Opt. Soc. Am. A*, 9(10):1749–1760, Oct 1992.

[8] J. A. Fessler and B. P. Sutton. Nonuniform fast Fourier transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560 – 574, 2003.

[9] J. Frank. *Three Dimensional Electron Microscopy of Macromolecular Assemblies*. Academic Press, Inc., 1996.

[10] J. Frank. Cryo-electron microscopy as an investigative tool: the ribosome as an example. *BioEssays*, 23(8):725–732, 2001.

[11] M. X. Goemans and D.P. Williamson. Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming. *Journal of the ACM*, 42:1115–1145, 1995.

[12] L. Greengard and J. Lee. Accelerating the Nonuniform Fast Fourier Transform. *SIAM Review*, 46(3):443–454, 2004.

[13] L. Lebart, A. Morineau, and K. M. Warwick. *Multivariate descriptive statistical analysis: correspondence analysis and related techniques for large matrices*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1984.

[14] G. Lerman, M. McCoy, J. A. Tropp, and T. Zhang. Robust computation of linear models, or How to find a needle in a haystack. `arXiv:1202.4044v1` [cs.IT], 2012.

[15] Z. Luo, W. Ma, A. So, Y. Ye, and S. Zhang. Semidefinite Relaxation of Quadratic Optimization Problems. *IEEE Signal Processing Magazine*, 27(3):20–34, may 2010.

[16] S.P. Mallick, S. Agarwal, D.J. Kriegman, S.J. Belongie, B. Carragher, and C.S. Potter. Structure and view estimation for tomographic reconstruction: A bayesian approach. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2253–2260, 2006.

[17] F. Mezzadri. How to generate random matrices from the classical compact groups. *Notices of the AMS*, 54:592–604, 2007.

[18] F. Natterer. *The Mathematics of Computerized Tomography*. Classics in Appl. Math. 32. SIAM, Philadelphia, 2001.

[19] H. Nyquist. Least orthogonal absolute deviations. *Computational Statistics & Data Analysis*, 6(4):361–367, June 1988.

[20] P. Penczek, R. Grassucci, and J. Frank. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3D cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3):251 – 270, 1994.

[21] P. Penczek, M. Radermacher, and J. Frank. Three-dimensional reconstruction of single particles embedded in ice. *Ultramicroscopy*, 40(1):33–53, 1992.

[22] P. A. Penczek, R. A. Grassucci, and J. Frank. The ribosome at improved resolution: New techniques for merging and orientation refinement in 3d cryo-electron microscopy of biological particles. *Ultramicroscopy*, 53(3):251 – 270, 1994.

[23] P. A. Penczek, J. Zhu, and J. Frank. A common-lines based method for determining orientations for N $>$ 3 particle projections simultaneously. *Ultramicroscopy*, 63(3-4):205 – 218, 1996.

[24] M. Radermacher, T. Wagenknecht, A. Verschoor, and J. Frank. A new 3-D reconstruction scheme applied to the 50S ribosomal subunit of E. coli. *Ultramicroscopy*, 141:RP1–2, 1986.

[25] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed Minimum-Rank Solutions of Linear Matrix Equations via Nuclear Norm Minimization. *SIAM Rev.*, 52(3):471–501, August 2010.

[26] I. I. Serysheva, E. V. Orlova, W. Chiu, M. B. Sherman, S. L. Hamilton, and M. van Heel. Electron cryomicroscopy and angular reconstitution used to visualize the skeletal muscle

calcium release channel. *Nat Struct Mol Biol*, 2:18–24, 1995.

[27] Y. Shkolnisky and A. Singer. Viewing direction estimation in cryo-EM using synchronization. *SIAM Journal on Imaging Sciences*, 5(3):1088–1110, 2012.

[28] A. Singer, R. R. Coifman, F. J. Sigworth, D. W. Chester, and Y. Shkolnisky. Detecting consistent common lines in cryo-EM by voting. *Journal of Structural Biology*, 169(3):312–322, 2010.

[29] A. Singer and Y. Shkolnisky. Three-Dimensional Structure Determination from Common Lines in Cryo-EM by Eigenvectors and Semidefinite Programming. *SIAM Journal on Imaging Sciences*, 4(2):543–572, 2011.

[30] A. Singer, Z. Zhao, Y. Shkolnisky, and R. Hadani. Viewing Angle Classification of Cryo-Electron Microscopy Images Using Eigenvectors. *SIAM Journal on Imaging Sciences*, 4(2):723–759, 2011.

[31] A. So, J. Zhang, and Y. Ye. On approximating complex quadratic optimization problems via semidefinite programming relaxations. *Math. Program.*, 110(1):93–110, March 2007.

[32] H. Späth and G. A. Watson. On orthogonal linear approximation. *Numer. Math.*, 51(5):531–543, October 1987.

[33] B. Vainshtein and A. Goncharov. Determination of the spatial orientation of arbitrarily arranged identical particles of an unknown structure from their projections. In *Proc. llth Intern. Congr. on Elec. Mirco.*, pages 459–460, 1986.

[34] M. van Heel. Multivariate statistical classification of noisy images (randomly oriented biological macromolecules). *Ultramicroscopy*, 13(1-2):165 – 183, 1984.

[35] M. van Heel. Angular reconstitution: A posteriori assignment of projection directions for 3D reconstruction. *Ultramicroscopy*, 21(2):111 – 123, 1987.

[36] M. van Heel and J. Frank. Use of multivariates statistics in analysing the images of biological macromolecules. *Ultramicroscopy*, 6(1):187 – 194, 1981.

[37] M. van Heel, B. Gowen, R. Matadeen, E. V. Orlova, R. Finn, T. Pape, D. Cohen, H. Stark, R. Schmidt, M. Schatz, and A. Patwardhan. Single-particle electron cryo-microscopy: towards atomic resolution. *Quarterly Reviews of Biophysics*, 33(04):307–369, 2000.

[38] L. Wang and F. J. Sigworth. Cryo-EM and single particles. *Physiology (Bethesda)*, 21:13–18, 2006.

[39] L. Wang and A. Singer. Exact and stable recovery of rotations for robust synchronization, 2012. submitted. Also availabe on Arxiv.

[40] Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2:203–230, 2010. 10.1007/s12532-010-0017-1.